ROOTS OF POLYNOMIALS EXPRESSED IN TERMS OF ORTHOGONAL POLYNOMIALS *

DAVID DAY † AND LOUIS ROMERO ‡

Abstract. A technique is presented for determining the roots of a polynomial $p(x)$ that is expressed in terms of an expansion in orthogonal polynomials. The roots are expressed as the eigenvalues of a nonstandard companion matrix $\mathbf{B}_n$ whose coefficients depend on the recurrence formula for the orthogonal polynomials, and on the coefficients of the orthogonal expansion. Some questions on the numerical stability of the eigenvalue problem to which they give rise are discussed. The problem of finding the roots of a transcendental function $f(x)$ can be reduced to the problem considered by approximating $f(x)$ by a Chebyshev polynomial. We illustrate the effectiveness of this convert-to-Chebyshev strategy by solving several transcendental equations using this plus our new algorithm. We analyze the numerical stability through both linear algebra theory and numerical experiments and find that this method is very well-conditioned.

Key words. rootfinding, Chebyshev polynomial, Legendre polynomial, single transcendental equation, global methods, companion matrix, eigenvalue problem

AMS subject classifications. 65H05, 42C10, 65H20, 65F15.

1. Introduction.

Suppose we want to find the real roots (especially those in $[-1,1]$) of a polynomial expressed by its Chebyshev coefficients,

$$p(x) = \sum_{i=0}^{n} \gamma_i T_i(x).$$

Or more generally, $p(x)$ may be expressed in terms of polynomials $\{\phi_m(x)\}_{m\geq 0}$, each $\phi_m(x)$ of exact degree m, that are orthogonal with respect to an inner product, e.g.

$$\langle f, g\rangle_\rho = \int_a^b f(x)\overline{g(x)}\rho(x)dx, \tag{1.1}$$

for some real and positive weight function $\rho(x)$.

One way to find the roots of $p(x)$ is to express $p(x)$ as a sum of monomials, and then to calculate the roots as the eigenvalues of the standard companion matrix. However, expressing a polynomial by its monomial coefficients is not as well conditioned as the expression in terms of Chebyshev polynomials. The transformation between a polynomial of degree n in $[-1,1]$ and its expansion coefficients with respect to the monomials [13] has $\mathcal{O}((1+\sqrt{2})^{n+1})$ condition number with respect to maximum norms (over $[-1,1]$) and with respect to Chebyshev polynomials [11] has $\mathcal{O}(n)$ condition number.

For the case of Chebyshev polynomials, Boyd [6] and also Battles and Trefethen [2] have proposed solving this problem by projecting to the unit circle in the complex z-plane with $x = (z+z^{-1})/2$, and using the fact that $T_k(x) = \cos(k\cos^{-1}(x))$. Their

*SANDIA IS A MULTIPROGRAM LABORATORY OPERATED BY SANDIA CORPORATION, A LOCKHEED MARTIN COMPANY, FOR THE US DEPARTMENT OF ENERGY'S NATIONAL NUCLEAR SECURITY ADMINISTRATION UNDER CONTRACT DE-AC04-94AL85000.

† Computational Mathematics and Algorithms Department, Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185-1110, (dmday@sandia.gov).

‡ Computational Mathematics and Algorithms Department, Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185-1110, (lromero@sandia.gov).

technique allows them to find the roots of $p(x)$ in terms of a standard companion matrix that depends on the coefficients $\{\gamma_k\}_{k=0}^n$ of the orthogonal expansion. These authors have found that this is a very successful algorithm, but the trouble is, it makes use of an eigenvalue problem of size $2n$ for a rootfinding problem of size n.

The present manuscript proposes an alternative formulation based on a nonstandard companion matrix $\mathbf{B}_n$ of dimension n. The algorithm is an extension of the technique [14] for finding the roots of the nth orthogonal polynomial $\phi_n(x)$. The technique uses the fact that any set of orthogonal polynomials satisfies a recurrence formula of the form

$$x\phi_{n-1}(x) = \sum_{i=0}^{n} \phi_i(x) h_{i,n-1}. \tag{1.2}$$

The coefficients determine an n by n matrix $\mathbf{H}_n = [h_{i,j}]_{0\le i,j<n}$ whose eigenvalues are the roots of the nth orthogonal polynomial $\phi_n(x)$. For ortho*normal* polynomials based on certain inner products such as Equation (1.1), $\mathbf{H}_n$ is symmetric and tridiagonal. For a general inner products, $\mathbf{H}_n$ is upper Hessenberg, that is, $h_{i,j} = 0$ for $i > j+1 > 0$.

As a specific example, the Chebyshev polynomials satisfy the three term recurrence $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$ for $n \ge 1$, or recast in the form of Equation (1.2), $xT_n(x) = T_{n-1}(x)\frac{1}{2} + T_{n+1}(x)\frac{1}{2}$. In this case, there holds $h_{0,1} = \frac{1}{2}$, $h_{0,1} = 1$, for $i > 0$, $h_{i,i+1} = h_{i+1,i} = \frac{1}{2}$ and otherwise $h_{i,j} = 0$. The asymmetry of $h_{0,1}$ and $h_{1,0}$ reflects the non-constant normalization of $\{T_k\}_{k\ge 0}$: for $\rho(x) = \sqrt{1-x^2}$ there holds $\langle T_k, T_k\rangle_\rho = \frac{\pi}{4}(1+\delta_{k,0})$.

Our technique for finding the roots of $p(x)$ is a modification of the technique for finding the roots of $\phi_n(x)$. To express our result we use the notation

$$\mathbf{f}_n(x) = [\phi_0(x), ..., \phi_{n-1}(x)]^T, \tag{1.3}$$

for the column vector-valued function containing the first n orthogonal polynomials, and the notation

$$\mathbf{c}^T = [\gamma_0, \gamma_1, ..\gamma_{n-1}], \tag{1.4}$$

for the column vector containing the first n coefficients of the polynomial $p(x)$. Using this notation we have

$$p(x) = \mathbf{f}_n(x)^T \mathbf{c} + \gamma_n \phi_n(x). \tag{1.5}$$

In §2, Theorem 2.3 shows that the roots of $p(x)$ are the eigenvalues of the nonstandard companion matrix

$$\mathbf{B}_n = \mathbf{H}_n - h_{n,n-1} \frac{\mathbf{c}}{\gamma_n} \mathbf{e}_{n-1}^T. \tag{1.6}$$

where $\mathbf{e}_{n-1} = [0, \ldots, 0, 1]^T$ is a column vector of dimension n. When applied to finding roots of polynomials expressed in terms of Chebyshev polynomials, for large values of n the new method promises to be something like eight times faster than the method proposed by Boyd and Battles and Trefethen. It is somewhat faster than the direct conversion to a monomials (without doubling the degree), which is unstable for large values of n.

Orthogonal polynomials have many applications. Transcendental equations may be solved with Chebyshev polynomials as was proposed in [4] and developed further in follow-up papers [5] and [6]. Battles and Trefethen automate, through MATLAB calls, a suite of operators on functions. The implementation is accomplished using Chebyshev polynomials of very high degree. And the operator that finds the real roots of a function is (now) implemented along the lines described here. Battles and Trefethen have pointed out that certain applications of polynomials based on the monomial form may be significantly improved by using another form based on a specific family of orthogonal polynomials.

Although the technique we present in this paper finds all of the roots of the polynomial $p(x)$, we will see that it only has desirable stability properties for finding roots in an appropriate region of the complex plane. For example, for Chebyshev polynomials we only have desirable stability properties for finding real roots in or near to the interval $[-1, 1]$. Similarly, transcendental equation solvers based on the rootfinding by Chebyshev expansions have desirable stability properties only for roots in or near to interval $[-1, 1]$ (see Theorem 4.2).

When we approximate a transcendental function in terms of an orthogonal polynomial expansion, the highest order coefficient γ_n converges to zero (see Theorem 4.1). For this reason, many cases of interest are near to the division by zero singularity in Equation (1.6) for $\mathbf{B}_n$. The singularity is avoided by solving a generalized eigenvalue problem as described in §2 or [20]. However, in Theorem 4.2, we will show that if a transcendental function is approximated as a finite sum of Jacobi polynomials, the roots found by using the corresponding matrix $\mathbf{B}_n$ accurately approximate the transcendental equation roots in or near $[-1, 1]$.

If the cost of solving the eigenvalue problem becomes a computational bottle neck, then one may use a subdivision algorithm (see [7] and [8]) that decomposes the rootfinding problem into several subproblems, and applies Chebyshev polynomials of lower order in each subinterval.

1.1. Summary. We begin in §2 by reviewing a process for finding the roots of the *nth* orthogonal polynomial $\phi_n(x)$ as the eigenvalues of the matrix $\mathbf{H}_n$. We then show how to modify this process to construct the nonstandard companion matrix $\mathbf{B}_n$ whose eigenvalues are given by the roots of the polynomial $p(x)$ (c.f. Theorem 2.3). Although classical orthogonal polynomials are emphasized over all, we abstractly define "orthogonal polynomial" (see Definition 2.1) so that our results include the monomials, and hence our results include the standard companion matrix. Lemma 2.4 presents analytical expressions for both the left and right eigenvectors in terms of the eigenvalues. In §3 the sensitivities of polynomial roots and matrix eigenvalues are compared. Theorem 3.3 demonstrates how eigenvalue and polynomial root sensitivities coincide in certain cases.

The algorithms presented herein are not so much new, as they are not widely known. The companion matrices for orthogonal polynomials were independently discovered by Hans Stetter. For a derivation of the nonstandard companion matrix based on quotient rings in algebraic geometry, see [22]. Exercise 1c on page 148 of [22] asks the reader to derive the companion matrix for Chebyshev polynomials. On the other hand, Stetter emphasizes application to polynomials of modest degree, say 10 (c.f. page 146). The observation that the roots of the nth member of a family of orthogonal polynomials must be the eigenvalues of a companion matrix whose elements come from the coefficients of the recurrence relation for the orthogonal polynomials was well known to C.J.G. Jacobi [14]. Like Stetter, we show how to define "orthogonal" polynomial broadly enough to apply the observation to any polynomial. Our contribution is some analysis of the numerical stability of such methods. Example 3 of §5 uses a degree 256 polynomal to solve a transcendental equation.

In order to concentrate on issues of interest in applications using orthogonal polynomials, we discuss the representative application of finding the roots of a scalar transcendental equation in a real interval. Representing a function by the partial sum of an exponentially convergent orthogonal expansion raises issues that must be addressed. In particular, ill-conditioning is manifested in the roots that we do not want. However, roots in a specific domain of the complex plane are well conditioned in a certain sense.

In §3.3 representation with respect to Chebyshev polynomials, or any Jacobi polynomial, are shown to be ideal for finding roots in or quite near to $[-1, 1]$. Away from $[-1, 1]$, the Jacobi polynomials are not recommended. The prerequisite results for classical orthogonal polynomials are reviewed. It is shown that for rootfinding in an interval, representing polynomials with respect to Jacobi orthogonal polynomials are ideal. But monomials are better for rootfinding in the unit disk. In particular the algorithms described herein are not designed to find all of the roots of a polynomial.

In §4 we discuss how matrix balancing is desirable in computing the eigenvalues of $\mathbf{B}_n$. The upper Hessenberg structure of $\mathbf{B}_n$ is crucial in the explanation of the success of matrix balancing. Theorem 4.2 shows how partial sums of orthogonal expansions lead to companion matrices that are amenable to matrix balancing. We use our analytical expressions for the left and the right eigenvalues to show that the polynomial and eigenvalue sensitivities differ by a computable (and benign) factor, related to the associated Lagrange interpolation polynomials. The companion matrix formulation is numerically stable in this case.

Analysis is also included intended for *a posteriori* use in solving transcendental equations. An algorithm for finding the roots of a transcendental equation in $[-1, 1]$ using expansions in terms of Chebyshev polynomials is presented in §4.2. Numerical experiments are presented in §5 that demonstrate the reliability of the algorithm. Our results are summarized in §6.

For expansions of transcendental equations, we explain why the companion matrix is amenable to balancing. The exponential convergence rate is related to the distance to the nearest singularity of the locally analytic function, and also applies to the (right) eigenfunctions. Balancing "factors out" the dependence of $\mathbf{B}_n$ on $1/\gamma_n$, and the balanced companion matrix eigenvalue problem is numerically stable. An algorithm for finding the roots of a transcendental equation on $[-1, 1]$ using expansions in terms of Chebyshev polynomials is presented in §4.2. Numerical experiments are presented in §5 that demonstrate the reliability of the algorithm.

2. Companion Matrices. Starting from a general definition of orthogonal polynomials, we review the procedure for finding the roots of orthogonal polynomials as the eigenvalues of the matrix $\mathbf{H}_n$ containing the coefficients in the recurrence formula. The discussion closely follows [14]. Next we construct a nonstandard companion matrix corresponding to a sequence of orthogonal polynomials and a given polynomial. In Theorem 2.3 we establish the equivalence between the roots of the polynomial equation, and the companion matrix spectrum. In Lemma 2.4 we give an analytical expression for the right eigenvectors of $\mathbf{B}_n$. The proof exploits the connection between Vandermonde matrices and Lagrange interpolation polynomials.

Orthogonal polynomials are broadly defined here to emphasize the connection between the numerical stability of a companion matrix eigenvalue problem and the associated inner product. There is a one to one correspondence between inner products on polynomials, and the set of sequences of univariate polynomials $\{\phi_i(x)\}_{i\geq 0}$ such that each $p_k(x)$ has degree k. The polynomials are orthonormal with respect to the polynomial inner product that is the ordinary vector inner product of expansion coefficients.

We will work over the space of complex valued continuous functions on a bounded subdomain of the complex plane.

DEFINITION 2.1. *With respect to the inner product* $\langle , \rangle$ *the sequence* $\{\phi_n(x)\}_{n\geq 0}$ *are* **orthogonal polynomials** *if each* $\phi_n(x)$ *is a polynomial of exact degree* n *and* $\langle \phi_n, \phi_m \rangle = \delta_{n,m}\sigma_n^2$. *Here* $\delta_{i,j}$ *is Kronecker's delta and* $\{\sigma_n\}_{n\geq 0}$ *is a sequence of positive real numbers. The polynomials* $\{\phi_n(x)\}_{n\geq 0}$ *are orthonormal if each* σ_n *is one. The norm induced by the inner product is denoted by* $\|\psi\| = \langle \psi, \psi \rangle^{1/2}$.

Orthogonality implies that for $i \leq n-1$, $\|\phi_i\|^2 \;\; h_{i,n-1} = \langle \phi_i(x), x\phi_{n-1}(x) \rangle$.

Usually when discussing orthogonal polynomials we will be concerned with inner products of the form in Equation (1.1). Orthonormal polynomials with respect to this type of inner product must satisfy a symmetric three term recurrence formula. This is a consequence of the fact that such an inner product is symmetric with respect to multiplication; that is, $\langle xf(x), g(x) \rangle = \langle f(x), xg(x) \rangle$.

For root finding problems over bounded complex domains, we recommend the inner product that arises in S. Bergman's theory of (reproducing) kernel functions (see [19] page 36 or [23] §11.2 or [18] Lemma 17.2.3). For example, the monomials are orthogonal polynomials with respect to the inner product

$$\langle f, g \rangle = \frac{1}{2\pi i} \int_{\Gamma} f(z)\overline{g}(z)dz,$$

where the integral is taken over the circle Γ centered around the origin in the complex plane. Note that this inner product is not symmetric with respect to multiplication.

2.1. Roots of Orthogonal Polynomials: A Review. A way to find the roots of the nth orthogonal polynomial $\phi_n(x)$ uses the recurrence formula in Equation (1.2). The first n instances of Equation (1.2) combine using the matrix $\mathbf{H}_n$, the column vector $\mathbf{f}_n$, and the coefficient $h_{n,n-1}$ into the matrix equation

$$x\mathbf{f}_n^T(x) = \mathbf{f}_n^T(x)\mathbf{H}_n + \phi_n(x)h_{n,n-1}\mathbf{e}_{n-1}^T. \tag{2.1}$$

Equation (2.1) exposes the equivalence between the roots ξ root of $\phi_n(x) = 0$ and the eigenvalues of $\mathbf{H}_n$,

$$\xi\mathbf{f}_n^T(\xi) = \mathbf{f}_n^T(\xi)\mathbf{H}_n.$$

We conclude with the following result, that W. Gautschi [14] attributes to Jacobi.

THEOREM 2.2. *The algebraic eigenvalues of* H_n *defined in Equation (2.1) coincide with the algebraic roots of the degree* n *orthogonal polynomial* $\phi_n(x)$.

Proof. The result is a corollary of Theorem 2.3. □

2.2. Nonstandard Companion Matrices. Assuming that $\gamma_n \neq 0$, we can use Equation (1.5) to express $\phi_n(x)$ as

$$\phi_n(x) = \frac{p(x) - \mathbf{f}_n(x)^T\mathbf{c}}{\gamma_n}. \tag{2.2}$$

If we substitute this expression for $\phi_n(x)$ into Equation (2.1) we arrive at the equation

$$x\mathbf{f}_n^T(x) = \mathbf{f}_n^T(x)\mathbf{H}_n + \frac{p(x) - \mathbf{f}_n(x)^T\mathbf{c}}{\gamma_n}h_{n,n-1}\mathbf{e}_{n-1}^T. \tag{2.3}$$

We now see that if ξ is a root of $p(x) = 0$, then

$$\xi \mathbf{f}_n^T(\xi) = \mathbf{f}_n(\xi)^T \mathbf{B}_n \tag{2.4}$$

where as in Equation (1.6)

$$\mathbf{B}_n = \mathbf{H}_n - h_{n,n-1} \frac{\mathbf{c}}{\gamma_n} \mathbf{e}_{n-1}^T.$$

This shows that if ξ is a root of $p(x)$ then it must be an eigenvalue of $\mathbf{B}_n$, with left eigenvector $\mathbf{f}_n^T(\xi)$. The converse is established in Theorem 2.3.

Equivalently, we could express the first n terms of our recurrence formula as

$$x\mathbf{f}_{n+1}(x)^T = \mathbf{f}_{n+1}^T(x) \begin{bmatrix} \mathbf{H}_n \\ h_{n,n-1}\mathbf{e}_{n-1}^T \end{bmatrix}.$$

When we combine this with the requirement that $p(x) = 0$ using Equation (1.5), we get the system of equations

$$\mathbf{f}_{n+1}^T(\xi) \begin{bmatrix} \mathbf{H}_n & \mathbf{c} \\ h_{n,n-1}\mathbf{e}_{n-1}^T & \gamma_n \end{bmatrix} = \xi \mathbf{f}_{n+1}^T(\xi) \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}. \tag{2.5}$$

Any root ξ of $p(x) = 0$ must be an eigenvalue of this generalized eigenvalue problem, with left eigenvector $\mathbf{f}_{n+1}(\xi)$.

As is the case for companion matrices, one may either solve the generalized eigenvalue problem in Equation (2.5) and discard an infinite eigenvalue, or find the eigenvalues as defined in Equation (2.4).

The standard backward stable algorithms for generalized and ordinary eigenvalue problems in Equations (2.5) and (2.4) are the QZ and QR algorithms respectively. Due to the upper Hessenberg form of these matrices, no initial transformation to Hessenberg form is required for either QZ or QR. For computing eigenvalues only in the average case, QR is three times faster than QZ [15]. The question of which formulation to use is not an entirely solved problem. The fact that $\mathbf{B}_n$ may have a very large norm suggests that the formulation of Equation (2.5) has superior stability properties. Numerical experiments do not confirm this hypothesis. Polynomial equations for which the formulation of Equation (2.5) is advantageous do exist [20], but do not arise in solution of transcendental equations. We performed numerical experiments comparing the residual norms of the polynomials evaluated at the eigenvalues computed by either QZ or QR. We observed that, if QR is used without balancing, then the eigenvalues computed from the ordinary eigenvalue problem suffer roundoff errors proportional to $\|\mathbf{c}\|/\gamma_n$. QR with balancing and QZ always computed eigenvalues of the same quality even if $\|\mathbf{c}\|/\gamma_n$ is very large. Explanations are provided in Theorem 3.3 and in §4.

Next the equivalence of the roots of the polynomial $p(x)$ and the eigenvalues of the matrix $\mathbf{B}_n$ defined in Equations (1.6) and (2.4) is demonstrated.

THEOREM 2.3. *The roots of a polynomial p of exact degree n coincide with the eigenvalues of the generalized companion matrix $\mathbf{B}_n$ counting algebraic multiplicity.*

Proof. We have already shown that a root ξ of $p(x)$ is an eigenvalue of $\mathbf{B}_n$ with left eigenvector $\mathbf{f}_n(\xi)$. The converse follows from two properties of unreduced upper Hessenberg matrices, including $\mathbf{B}_n - \xi I_n$ for any ξ: first a nontrivial (right) null vector has nonzero last component, and second the nullity is at most one. The properties of Hessenberg matrices are discussed in [15] §7.4.5 and in particular Theorem 7.4.4.

If ξ is an eigenvalue of $\mathbf{H}_n$ with nontrivial (right) eigenvector $\mathbf{v} = [v_0, ...v_{n-1}]^T$, then $v_{n-1} \neq 0$. Substitution of Equation (2.2) into Equation (2.1) yields

$$x\mathbf{f}_n^T(x) = \mathbf{f}_n^T(x)\left(\mathbf{H}_n - \frac{\mathbf{c}}{\gamma_n}h_{n,n-1}\mathbf{e}_{n-1}^T\right) + \frac{p(x)}{\gamma_n}h_{n,n-1}\mathbf{e}_{n-1}^T. \tag{2.6}$$

Inspection of the product of Equation (2.6) and $\mathbf{v}$ implies that $p(\xi) = 0$. As a consequence of the second property, a left eigenvector of ξ must be proportional to $\mathbf{f}_n(\xi)$. ◻

2.3. The Right Eigenvectors. We have already shown that if ξ_j is the jth root of the polynomial $p(x)$, then $\mathbf{v}_j^T = \mathbf{f}_n^T(\xi_j)$ is the left eigenvector associated with the eigenvalue ξ_j of $\mathbf{B}_n$. We will now give a simple expression for the right eigenvectors $\mathbf{w}_j$ associated with this eigenvalue.

The matrix of left eigenvectors, $\mathbf{V}$, has jth row $\mathbf{v}_j$. The ith row of the inverse contains the right eigenvector $\mathbf{w}_i$. Note that the n by n matrix $\mathbf{V} = [\nu_{i,j}]$, is called a generalized Vandermonde matrix due to $\nu_{i,j} = \phi_j(\xi_i)$.

The right eigenvectors can be expressed using the interpolating polynomials. Assuming that ξ_j is a simple root of $p(x)$, the jth Lagrange interpolating polynomial associated with the roots of $p(x)$ is

$$l_j(x) = \frac{p(x)}{p'(\xi_j)(x - \xi_j)}.$$

Each interpolating polynomial $l_j(x)$ has degree $n - 1$ and satisfies $l_j(\xi_k) = \delta_{jk}$ for $0 \leq j, k < n$.

Each polynomial $l_j(x)$ has degree $n-1$, and is a linear combination of $\{\phi_k(x)\}_{k=0}^n$. Define the column vector $\mathbf{w}_j$ to contain the expansion coefficients of $l_j(x)$,

$$l_j(x) = \mathbf{f}_n^T(x)\mathbf{w}_j.$$

It follows that $\delta_{ij} = l_j(\xi_i) = \mathbf{f}_n^T(\xi_i)\mathbf{w}_j = \mathbf{v}_i^T\mathbf{w}_j$. This shows that the vector $\mathbf{w}_i$ is in fact the ith column of the inverse matrix of $\mathbf{V}$, and hence $\mathbf{w}_i$ is the right eigenvector associated with the eigenvalue ξ_i. This proves the following lemma.

LEMMA 2.4. *The companion matrix $\mathbf{B}_n$ defined in Equation (2.4) corresponding to a polynomial $p(x)$ of exact degree n and with distinct roots, $\{\xi_j\}_{0\leq j<n}$, has as eigenvalues the roots of $p(x)$. The left eigenvector corresponding to ξ_k is $\mathbf{f}_n(\xi_k)$. Moreover, if $\{\phi_j\}$ are orthonormal polynomials, then for $l_j(x) = \frac{p(x)}{p'(\xi_j)(x-\xi_j)}$ the jth right eigenvector $\mathbf{w}_j$ has components $\mathbf{e}_k^T\mathbf{w}_j = \langle l_j, \phi_k\rangle$, and $\|\mathbf{w}_j\|_2 = |\!|\!| l_j |\!|\!|$.*

Proof. The comments directly before the Lemma establish the following: the n by n matrix $\mathbf{W} = [\mathbf{w}_0, ..., \mathbf{w}_{n-1}]$ whose jth column, $\mathbf{w}_j$, contains the expansion coefficients of $l_j(x)$ with respect to $\{\phi_j(x)\}_{0\leq j<n}$, i.e. $l_j(x) = \mathbf{f}_n^T(x)w_j$, is $\mathbf{W} = \mathbf{V}^{-1}$, thus is the matrix of right eigenvectors. Next take the inner product of $l_j(x)$ with $\phi_k(x)$. By Definition 2.1, $\langle l_j, \phi_k\rangle = |\!|\!|\phi_k|\!|\!|^2 e_k^T w_j$. The desired representation follows in the orthonormal case. Parseval's formula readily furnishes the equivalence between the norms. ◻

The relationship between $|\!|\!| l_j |\!|\!|$ and $p(x)$ is further developed in §4.

3. Sensitivity Analysis. Root finding by eigenvalue problems is popular due to its favorable stability properties compared to other methods. A stability analysis for the standard companion matrix formulation has been performed in [24]. The companion matrix form is viewed as a rank one perturbation of a bidiagonal matrix, and the inverses of the shifted companion matrices are analyzed. Here we perform a local analysis of the nonstandard companion matrix, assessing stability by comparing the polynomial root and eigenvalue the polynomial root sensitivities. The limitation of a local analysis is that it is only valid for eigenvalues that are well separated.

3.1. Polynomial Root Sensitivity. The perturbation theory for polynomial roots is considered following [12]. We consider some particular zero ξ of $p(x)$ as a complex valued function of the expansion coefficients with respect to the orthogonal polynomials. Denote the dependence on the coefficients $\mathbf{c}$ by $\xi(\mathbf{c})$.

LEMMA 3.1. *Suppose that a polynomial of exact degree n, $p(x) = \mathbf{f}_{n+1}^T(x)\mathbf{c}_o$, where $\mathbf{c}_o$ is a $n+1$ dimensional column vector, has a simple root ξ_o, such that $p(\xi_o) = 0 \neq p^{'}(\xi_o)$. There exists a smooth function $\xi(\mathbf{c})$ such that $\xi(\mathbf{c}_o) = \xi_o$ and $\mathbf{f}_{n+1}^T(\xi(\mathbf{c}))\mathbf{c} = 0$, with gradient $\nabla_{\mathbf{c}}\xi(\mathbf{c}_o) = -\mathbf{f}_{n+1}^T(\xi_o)/p^{'}(\xi_o)$.*

Proof. See Example 3.10 in [22] □

We will discuss several aspects of Lemma 3.1, starting with the excluded case of multiple roots. In the case of a root of algebraic multiplicity $m > 1$, there exist infinitesimal coefficient perturbations that change the multiplicity, and the roots are Hölder continuous with exponent $1/m$ (see Proposition 5.1 in [22] or Theorem 4.1 in [6]). Approximations to roots with nontrivial multiplicity correspond to small values of $p(\xi_o)$. Monitoring the value of the polynomial derivative at all approximate roots of interest for small values is required.

Next, the implication of Lemma 3.1 is that the norm of the gradient of a root with respect to the coefficients is proportional to $\|\mathbf{f}_{n+1}(\xi_o)\|_2$. In words, a root ξ_o is not very sensitive to the polynomial coefficients if both $\|\mathbf{f}_{n+1}(\xi_o)\|_2$ is not "large", and $|p'(\xi_o)|$ is not "small". For all the orthogonal polynomials familiar to the authors, for sufficiently large x, $\|\mathbf{f}_{n+1}(x)\|_2$ grows exponentially as a function of n. This observation reflects the intrinsic difficulty of finding all the roots of an arbitrary polynomial. On the other hand, using Jacobi polynomial, for a root $-1 \leq \xi_o \leq 1$, $\|\mathbf{f}_{n+1}(\xi_o)\|_2$ is not "large" (clarified in §3.3) and no well separated root is very sensitive to the coefficients.

3.2. Eigenvalue Sensitivity. In §3.1 we showed that the polynomial root sensitivity with respect to the coefficients is $\|\mathbf{f}_{n+1}\|$. Here the condition number of a simple eigenvalue is related to the corresponding left and right eigenvectors $\mathbf{f}_n = \mathbf{v}$ and $\mathbf{w}$. A standard result from the perturbation theory of simple eigenvalues is that under an infinitesimal perturbation $\delta\mathbf{A}$ of a matrix $\mathbf{A}$, a simple eigenvalue λ changes to $\lambda + \delta\lambda$ where $\delta\lambda = \mathbf{v}^T\delta\mathbf{A}\mathbf{w}\ /\mathbf{v}^T\mathbf{w}$. Lemma 3.2 restates the result without using infinitesimals.

LEMMA 3.2. *If a square matrix $\mathbf{A}$ has a simple eigenvalue λ with corresponding left $\mathbf{v}^T$ and right $\mathbf{w}$ eigenvectors, $\mathbf{v}^T\mathbf{A} = \lambda\mathbf{v}^T$ and $\mathbf{A}\mathbf{w} = \mathbf{w}\lambda$, then $\nabla_{\mathbf{A}}\lambda = \mathbf{v}\mathbf{w}^T/\mathbf{v}^T\mathbf{w}$.*

Proof. See [15] page 344. □

Next Theorem 3.3 gives a sufficient condition for the numerical stability of root-finding based on a nonstandard companion matrix eigenvalue problem. It suffices for the computed eigenvalues to be the eigenvalues of $\mathbf{B}_n + \mathbf{E}_n$ nearby to $\mathbf{B}_n$ in a component-wise or relative sense. That is, there exists a tiny $\tau > 0$ such that $|\mathbf{E}_n| \leq \tau|\mathbf{B}_n|$. The result applies for any nonzero γ_n. The idea of the proof is that the factor of $1/\gamma_n$ in column n of $\mathbf{E}_n$ cancels with a factor of γ_n in row n of $\mathbf{w}$.

THEOREM 3.3. *Suppose that the degree n polynomial $p(x)$ has a simple root λ. Recall the notation of Equation (1.5), in particular the definition of the column $\mathbf{c}$ in Equation (1.4), and the definition of $\mathbf{H}_n$ in Equation (2.1). The corresponding com-*

panion matrix $\mathbf{B}_n$ *has left and right eigenvectors* $\mathbf{v}^T = \mathbf{f}_n(\lambda)^T$ *and* $\mathbf{w}$ *as in Lemma 2.4 so that* $\mathbf{v}^T\mathbf{w} = 1$*. A perturbation* $\mathbf{E}_n$ *of* $\mathbf{B}_n$ *such that* $|\mathbf{E}_n| \le \tau|\mathbf{B}_n|$ *perturbs* λ *by* $\delta\lambda$ *such that*

$$|\delta\lambda| \le \tau|\mathbf{v}|^T|\mathbf{H}_n|\;|\mathbf{w}| + \tau|\mathbf{v}|^T|\mathbf{c}|/|p'(\lambda)| + \mathcal{O}(\tau^2).$$

Proof. We start by establishing the following claim. For $h_{n,n-1}$ defined in Equation (1.2) there holds

$$h_{n,n-1}\mathbf{e}_{n-1}^T\mathbf{w} = \gamma_n/p'(\lambda). \tag{3.1}$$

Substitute the expansion $l(x) = \mathbf{f}_n(x)\mathbf{w}$ below, and simplify to find that $\gamma_n = \langle p, \phi_n\rangle = \langle l(x)(x-\lambda)p'(\lambda), \phi_n\rangle = \langle l(x)xp'(\lambda), \phi_n\rangle = \langle \phi_{n-1}(x)xp'(\lambda), \phi_n\rangle\mathbf{e}_{n-1}^T\mathbf{w}$. Rewrite Equation (1.2) in matrix form,

$$x\phi_{n-1}(x) = \mathbf{f}_n^T(x)\mathbf{H}_n e_{n-1} + \phi_n(x)h_{n,n-1}. \tag{3.2}$$

Equation (3.2) implies that $\gamma_n = h_{n,n-1}p'(\lambda)\mathbf{e}_{n-1}^T\mathbf{w}$. Divide by $p'(\lambda)$ to establish the claim.

By Lemma 3.2, to first order in τ there holds $|\delta\lambda| = |\mathbf{v}^T\mathbf{E}_n\mathbf{w}| \le |\mathbf{v}|^T|\mathbf{E}_n||\mathbf{w}| \le \tau|\mathbf{v}|^T|\mathbf{B}_n||\mathbf{w}|$. The proof is completed by substituting Equation (1.6), applying the triangle inequality, and then the claim. □

Theorem 4.2 will show how for transcendental equations, QR iteration with balancing solves a nearby eigenvalue problem, $\mathbf{B}_n + \mathbf{E}_n$, such that only the last column of $\mathbf{E}_n$ is proportional to $1/\gamma_n$.

3.3. Classical Orthogonal Polynomials. In general, the term $\|\mathbf{f}_{n+1}(\xi)\|_2$ arising in polynomial roots sensitivity is associated with the "kernel polynomials". The kernel polynomials are defined by $\mathbf{K}_n(x_o, x) = \overline{\mathbf{f}}_{n+1}^T(x_o)\;\mathbf{f}_{n+1}(x)$. If x_o is a constant, then $\mathbf{K}_n(x_o, x)$ is a polynomial. The kernel polynomials maximize the ratio $|p(x_o)|/|\!|\!|p(x)|\!|\!|$ over all polynomials of exact degree n (see [23] Theorem 3.1.3), and the maximum ratio is $\|\mathbf{f}_{n+1}(x_o)\|_2$. As we shall see, the asymptotic properties of the kernel polynomials indicates that the classical orthogonal polynomials over $[-1, 1]$, the Jacobi polynomials, are suitable for rootfinding. And conversely, for rootfinding over domains that are topologically different from intervals, none of the classical polynomials orthogonal over an interval is desirable, and a different inner product is needed [1, 10].

The Jacobi polynomials are orthogonal with respect to the inner product

$$\langle f, g\rangle_{\alpha,\beta} = \int_{-1}^{1} f(t)g(t)(1-t)^\alpha(1+t)^\beta dt$$

for $\alpha > -1, \beta > -1$ (see [23] §2.4). The case $\alpha = \beta = -1/2$ corresponds to the Chebyshev polynomials (of the first kind). The comparison of spectral methods for partial differential equations based on either Chebyshev polynomials or Legendre polynomials ($\alpha = \beta = 0$) in [3] demonstrates the advantages of Legendre polynomials. The Legendre orthonormal polynomials satisfy the three term recurrence $\gamma_{n+1}\phi_{n+1}(x) = x\phi_n(x) - \gamma_n\phi_{n-1}(x)$ for $\gamma_n = n/\sqrt{4n^2-1}$. By rearranging the three term recurrence into the form of Equation (1.2), one may show that for $i \ge 0$, $h_{i+1,i} = h_{i,i+1} = \gamma_{i+1}$ and otherwise $h_{i,j} = 0$.

Theorem 7.71.2 in reference [23] states that $\max_{-1\le x_o\le 1}\|\mathbf{f}_{n+1}(x_o)\|_2 = \mathcal{O}(n^\kappa)$ for $\kappa = \max(\alpha+1, 1/2)$. The result is a consequence of the connection to Sturm-Liouville problems. The exponent is minimal for the Chebyshev polynomials, and for any admissible (α,β) the sensitivity bound grows like a polynomial. On the other hand, away from the interval $[-1,1]$, the Jacobi polynomials grow exponentially with degree, and are undesirable for root finding problems.

4. Solving Transcendental Equations. We will discuss in depth a representative application of nonstandard companion matrices to rootfinding problems involving polynomial equations expressed with respect to Chebyshev polynomials. We discuss the solution of a transcendental equation $\psi(\xi) = 0$. §4.2 presents a complete description of the corresponding algorithm for a scalar transcendental equation. In numerical experiments, we find that the algorithm is numerically stable if matrix balancing is used (the default in MATLAB). However it is crucial to use balancing with the QR algorithm in solving the eigenvalue problems that arise in the solution of transcendental equations.

4.1. Analysis of Balancing. Next some of the linear algebra issues associated with computed the eigenvalues of the B matrices are discussed in detail. Readers only interested in the solution of transcendental equations may choose to skip the section.

As a sequence of polynomials converge uniformly to $\psi(x)$ on some bounded domain, certain roots of the polynomials converge to $\{\xi : \psi(\xi) = 0\}$ [5]. The order of the approximation, n, is chosen to be sufficiently large that the trailing γ_n is negligible [5]. For solving transcendental equations, balancing the generalized companion matrix (c.f. [24]) usually employs an alarmingly ill conditioned diagonal similarity transform, and extraordinarily reduces the condition number of the matrix of eigenvectors. Theorem 4.2 presents an explanation of the success of balancing for companion matrices arising in the solution of transcendental equations.

A transcendental equation $\psi(\zeta) = 0$ arises from $\psi(x)$ that is analytic in an open simply connected domain containing $[-1,1]$. The orthogonal polynomials used are eigenfunctions of singular Sturm-Liouville problems in $[-1,1]$, namely the Jacobi polynomials (corresponding to one value of (α,β)) and here denoted $\{\phi_n(x)\}_{n\ge 0}$. The simplest case, $\{\phi_n(x)\}_{n\ge 0}$ orthonormal, is discussed. The convergence properties of Jacobi series expansion $\sum_{n\ge 0}\phi_n(x)\gamma_n$ with $\gamma_n = \langle\psi,\phi_n\rangle$ of $\psi(x)$ is described by Theorem 4.1.

THEOREM 4.1. *Let $\psi(x)$ be an analytic function with an open domain containing $[-1,1]$. The expansion of $\psi(x)$ in a Jacobi series is convergent in the interior of the greatest ellipse with foci at ± 1, in which $\psi(x)$ is regular. The expansion is divergent in the exterior of the ellipse. If $\psi(x) = \sum_{n\ge 0}\gamma_n\phi_n$, then we have the following representation of the sum R of the semi-axes of ellipse of convergence $R = \liminf_{n\to+\infty}|\gamma_n|^{-1/n}$.*

Proof. See [23] Theorem 9.1.1. □

In Theorem 4.1, $R = A+B$ for an ellipse $(x/A)^2 + (y/B)^2 = 1$ in the (x,y) plane with $A^2 = B^2 + 1$ and $A > B > 0$ (see [16] p.37). Roughly speaking, there holds $\sum_{j\ge n}|\gamma_j|^2 = \mathcal{O}(R^{-2n})$. The hypothesis that $\psi(x)$ is an analytic function on an open domain containing $[-1,1]$ ensures that for the greatest ellipse $B > 0$ and $R > 1$.

An analytic function $\psi(x)$ with root ξ, $\psi(\xi) = 0$, has a Jacobi series. The Jacobi series has partial sums of the form $p_n(x) = \sum_{j=0}^{n}\phi_j(x)\gamma_j$ Each $p_n(x)$ has at least one root root ξ_n nearest to ξ. In the case $R > 1$, Theorem 4.1 implies that in $[-1,1]$, $\{p_n(x)\}_{n\ge 0}$ converges uniformly to $\psi(x)$. Furthermore, each derivative $p_n^{(m)}(x)$

converges uniformly to $\psi^{(m)}(x)$ in $[-1,1]$.

If ξ is a simple root of $\psi(x)$ (i.e. $\psi'(\xi) \neq 0$), then $\xi_n \to \xi$. Moreover, for n sufficiently large that $\psi'(\xi_n)\psi'(\xi) > (\psi'(\xi))^2/2$ the relationship between residual error and approximate solution error implies that $\xi_n - \xi = \mathcal{O}(R^{-n})$.

The algebraic eigenvalue problem $\mathbf{B}_n\mathbf{W}_n = \mathbf{W}_n\Lambda_n$ is solved by applying the QR algorithm to the balanced generalized companion matrix. In MATLAB, the default configuration of the QR algorithm applies balancing. Balancing refers to determining a diagonal matrix $\mathbf{\Sigma}_n$ such that the similar eigenvalue problem $\mathbf{\Sigma}_n^{-1}\mathbf{B}_n\mathbf{\Sigma}_n$ is (hopefully) much better conditioned. A nearly optimal diagonal similarity transformations $\mathbf{\Sigma}_n$ produces $\mathbf{\Sigma}_n^{-1}\mathbf{W}_n$ with equal row norms (see [17] §12), but $\mathbf{W}_n$ is not known *a priori*. Instead a diagonal similarity transformation $\mathbf{\Sigma}_n$ that nearly minimizes a norm of $\mathbf{\Sigma}_n^{-1}\mathbf{B}_n\mathbf{\Sigma}_n$ is determined.

To illustrate matrix balancing consider $\mathbf{B}_4$ whose coefficients are chosen to reflect the asymptotic equation $\gamma_k = \mathcal{O}(R^{-k})$ for $R > 1$. A rootfinding algorithm based on Chebyshev polynomials is used. We use slightly more complicated coefficients, $\mathbf{c}^T = [2, 2R^{-1}, 2R^{-2} + \frac{1}{2}R^{-4}, 2R^{-3}]$ and $\gamma_4 = -R^4$, so that $\mathbf{B}_4$ takes the simple form in Equation (4.1). We have included an extra nonzero element in the south-west term to illustrate the essential contribution of the upper Hessenberg structure of $\mathbf{B}_n$ to the success of the matrix balancing algorithm. We approximately balance this matrix using $\mathbf{\Sigma}_4 = \mathrm{diag}(R^3, R^2, R^1, 1)$,

$$\mathbf{\Sigma}_4^{-1}\begin{bmatrix} 0 & 1/2 & 0 & R^4 \\ 1 & 0 & 1/2 & R^3 \\ 0 & 1/2 & 0 & R^2 \\ S & 0 & 0 & R \end{bmatrix}\mathbf{\Sigma}_4 = \begin{bmatrix} 0 & \frac{1}{2R} & 0 & R \\ R & 0 & \frac{1}{2R} & R \\ 0 & R/2 & & R \\ SR^3 & 0 & 0 & R \end{bmatrix}. \tag{4.1}$$

Note that because $\mathbf{B}_4$ is upper Hessenberg, $S = 0$, so that balancing reduces $\mathbf{B}_4$ in norm from $\mathcal{O}(R^4)$ to $\mathcal{O}(R)$. In this example, as R increases, the diagonal matrix determined by the balancing algorithm converges to $\mathbf{\Sigma}_4$. For polynomials of degree n, the norm of $\mathbf{B}_n$ is proportional to R^k for k possibly as large as n.

Different normalizations of the orthogonal polynomials correspond to the different diagonal similarity transformations applied to $\mathbf{B}_n$. The product of Equation (2.3) and $\mathbf{\Sigma}_n = \mathrm{diag}(\sigma_0, ..., \sigma_{n-1})$ has the form

$$x \ \ \mathbf{f}_n^T(x)\mathbf{\Sigma}_n = \mathbf{f}_n^T(x)\mathbf{\Sigma}_n \ \ \mathbf{\Sigma}_n^{-1}\mathbf{B}_n\mathbf{\Sigma}_n + \frac{p(x)}{\gamma_n}h_{n,n-1}\sigma_{n-1}\mathbf{e}_{n-1}^T. \tag{4.2}$$

In this sense, the balancing algorithm determines a suitable normalization of the orthogonal polynomials (c.f. Definition 2.1). Bear in mind that $e_j^T\mathbf{B}_n e_{n-1}$ is proportional to γ_j/γ_n.

The next Theorem will show that asymptotically the right eigenvectors all are graded in exactly the same way, decreasing from term to term by a ratio of approximately $1/R$. For such $\mathbf{B}_n$ for an optimal $\mathbf{\Sigma}_n$, which approximately equalizes the row norms of $\mathbf{\Sigma}_n^{-1}\mathbf{W}_n$, σ_i/σ_i is asymptotically R. In general, it is not necessarily the case that the diagonal $\mathbf{\Sigma}_n$ that approximately minimize a norm of $\mathbf{\Sigma}_n^{-1}\mathbf{B}_n\mathbf{\Sigma}_n$ is nearly optimal for the eigenvalue problem. For transcendental equation solving, asymptotically the polynomial coefficients also decrease by a factor of $1/R$ from coefficient to coefficient. Equation (4.1) illustrates how in this case the balancing algorithm determines a nearly optimal scaling for eigenvalue problems.

By Theorem 4.1, $(\gamma_n)_{n\geq 0}$ decays exponentially. Not surprisingly, in practice the diagonal elements of $\mathbf{\Sigma}_n$ exhibits similar exponential decay. The resulting $\mathbf{\Sigma}_n$ has an alarmingly large condition number. Next Theorem 4.2 will show that the rows of $(\mathbf{e}_j^T \mathbf{W}_n \mathbf{e}_k)_{0 \le j < n}$ decay at the same exponential rate as $(\gamma_n)_{n \ge 0}$. The transformation from $\mathbf{W}_n$ to $\mathbf{\Sigma}_n^{-1}\mathbf{W}_n$ reduces the variation in the norms of the rows of $\mathbf{W}_n$, and improves the condition number of the eigenvalue problem.

THEOREM 4.2. *Suppose that $\psi(x)$ is analytic in an ellipse with foci at ± 1, and that ξ is a simple root of $\psi(x)$ within the ellipse. Suppose in addition that for each partial sum of the Jacobi series expansion of $\psi(x)$, $p_n(x)$, all of the roots of $p_n(x)$ are within the ellipse. Choose a root ξ_n of $p_n(x)$ nearest to ξ. The generalized companion matrix corresponding to $p_n(x)$, $\mathbf{B}_n$, has an eigenvector $\mathbf{w}_n$ such that $\mathbf{B}_n \mathbf{w}_n = \mathbf{w}_n \xi_n$. Then $\mathbf{e}_j^T \mathbf{w}_n = \mathcal{O}(R^{-n})$.*

Proof. The maximal ellipses for $l(t) = (\psi(t) - \psi(\xi))/(t - \xi)$ and $\psi(t)$ coincide. For $l(t) = \sum_{n \ge 0} \phi_n(t)\mu_n$, by Theorem 4.1, there holds $R = \liminf |\mu_n|^{-1/n}$. The partial sums are $l_n(x) = \sum_{0 \le j \le n} \phi_j(x)\mu_j$. By careful accounting, one may show that for each fixed $\epsilon > 0$, and for $|x - \xi| > \epsilon$, there holds $l_n(x) - l(x) = \mathcal{O}(R^{-n})$. Furthermore a similar argument shows that $l_n(\xi_n) - l(\xi_n) = \mathcal{O}(R^{-n})$, from which $\| l_n(x) - l(x) \| = \mathcal{O}(R^{-n})$ follows. By Theorem

$$|\mathbf{e}_j^T \mathbf{w}_n| = |\langle \phi_j, l_n \rangle| = |\langle \phi_j, l_n - l \rangle + \langle \phi_j, l \rangle| = |\langle \phi_j, l_n - l \rangle + \mu_j| \le \| l_n - l \| + |\mu_j| = \mathcal{O}(R^{-n}).$$

□

Note that if the Jacobi series converges super exponentially, or even if R is very large, our justification of the balancing algorithm breaks down. We performed many numerical experiments, in floating point arithmetic with machine precision 2^{-54}, attempting to cause the balancing algorithm to fail. The expansion coefficients in the Jacobi series of a transcendental function coefficients converge to zero. We assume that each expansion coefficient, γ_m, with the maximal absolute value, $\sup_k |\gamma_k| = |\gamma_m|$, arises for $m \ll n$. For an entire function, $\lim_{n \to +\infty} \gamma_n/\gamma_{n+1} = +\infty$. The values of $\{\gamma_n\}_{n \ge 0}$ computed in finite precision arithmetic do not share this asymptotic property. The absolute error in each nonzero γ_n is, very roughly, the product of the machine precision and $\sup_k |\gamma_k|$. In our numerical experiments, we never observed a huge value of γ_n/γ_{n+1} for the nonzero approximate values of $\{\gamma_n\}_{n \ge 0}$. In other words, although matrix balancing has always worked for us, one must check that balancing determines a B_n not much larger in norm than H_n.

4.2. An Algorithm for Transcendental Equations. An algorithm is implemented as a MATLAB script that approximates the roots in an interval of a transcendental equation. Modified companion matrices are used to find the roots in or very near to $[-1, 1]$. The case of a polynomial expressed as a sum of Chebyshev polynomials is considered. The algorithm to approximate a function by a polynomial is reviewed briefly. Many subtle numerical analysis issues are discussed that are crucial for readers who actually want to solve a transcendental equation, such as rules for when to discard some of the eigenvalues. Readers more interested in concrete information on how to find the roots of of a given polynomial expressed as a sum of Chebyshev coefficients are directed to the paragraph directly following the algorithm.

A collocation method is used to determine the expansion coefficients with respect to the Chebyshev polynomials of a polynomial approximation of a scalar function $\psi(x)$ whose domain includes $[-1, 1]$ (see Appendix A in [6]). For completeness, we briefly review the popular method here. The Chebyshev Gauss Lobatto (CGL) points, $\cos(k\pi/n)_{k=0}^n$, are unisolvent. A unique nth degree polynomial interpolates $\psi(x)$ at the CGL points. The column vector of expansion coefficients $[\gamma_0, ..., \gamma_n]^T$ is the product of discrete Chebyshev transformation matrix, Π_n, and the column vector

$[\psi(\cos(0\pi/n)), ..., \psi(\cos(n\pi/n)]^T$, where $\Pi_n = [\cos(ij\pi/n)2/(q_iq_jn)]_{0\le i,j\le n}$, and $q_0 = q_n = 2$ and $q_i = 1$ otherwise. Other issues including spectral convergence, the adaptive selection of n, and the subdivision of the interval are discussed elsewhere [6].

Techniques for discarding some of the eigenvalues are discussed. There are two reasons to discard certain computed eigenvalues. First equations with n_p roots in or near to $[-1, 1]$ may be approximated by a polynomial of higher degree $n > n_p$. In finite precision arithmetic, the $n - n_p$ additional eigenvalues do not necessarily solve the polynomial equation. We would like to be able to reliably determine n_p. Second, in many applications the cost of evaluating the function is significant.

An important application of Chebyshev polynomials is solving transcendental equations in a way that minimizes the number of function evaluations [4]. For polynomials of high degree, some definitions of nearness to $[-1, 1]$ will classify a large percentage of the eigenvalues as potential polynomial roots, significantly increasing the number of function evaluations needed for equation residuals. The spurious eigenvalues are in a region in the complex plane in which Chebyshev polynomials of a given degree are wildly unstable. We only select eigenvalues within a domain of interest; here we discard eigenvalues outside of $(-2, 2) \times (-.2, .2)$. On the other hand, discarding roots may also be discarding part of the answer. Real roots may be approximated by complex eigenvalues near to [-1,1]. For example if the complex QR algorithm is applied to the real matrix $\mathbf{B}_n$ (for robustness), the set of computed roots is not closed under conjugation. The problem is addressed by using a partial condition number of the eigenvalues. The condition number of an eigenvalue, ξ, is the product of two terms, $\|\mathbf{f}_n(\xi)\|_2$ (defined in Equation (1.3)) and a term that involves the Lagrange interpolation polynomial whose support contains the eigenvalue. Our solution is to add a test that discards ξ such that $\|\mathbf{f}_n(\xi)\|_2$ is enormous (compare the definition of `cond_max`). Chebyshev polynomials are wildly unstable in such regions in the complex plane. At a multiple roots near to $[-1, 1]$, the norm of the vector values of the orthogonal polynomial evaluated at the roots is of order one, and is not discarded.

The parameters are chosen here to avoid large numbers of spurious roots. No attempt is made to find all of the roots of the polynomial.

```
n              = 2^4;                            % polynomial degree
[CGLpoints, ChebTransMat] = setupChebyshev(n);
FunctValues    = problemRod(CGLpoints);          % evaluate @ CGL pts
ExpansionCoeff = FunctValues * ChebTransMat;
if ExpansionCoeff(n+1) == 0,
   error('leading expansion coefficient vanishes; try --n');
end
ExpansionCoeff = ExpansionCoeff/(-2*ExpansionCoeff(n+1)); % normalize
H = diag(ones(n-1, 1)/2, 1) + diag(ones(n-1, 1)/2, -1); H(1, 2) = 1;
C = H; C(n, :) = C(n, :) + ExpansionCoeff(1:n); % nonstandard
Eigenvalues    = eig( C );                       % ... companion matrix
Vandermonde    = evalCheb(n,Eigenvalues);        % generalized ...
Vcolsums       = sum( abs(Vandermonde') );       % Vandermonde matrix
tube_index     = find((abs(imag(Eigenvalues))<.2) & ...
                      (abs(real(Eigenvalues))< 2));
Solutions      = Eigenvalues( tube_index );      % Spectrum in
Vcolsums       = Vcolsums( tube_index );         % ...(-2,2)x(-.2,.2)
cond_max       = min( 2^(n/2), 10^6 );           % Cluster threshold
condEigs_index = find ( Vcolsums < cond_max );   % Select
```

```
Solutions      = Solutions( condEigs_index );   % ... roots
```

Suppose we want to find the roots of the polynomial

$$p(x) = eT_0(x) + 2\pi T_1(x) + 2\gamma T_2(x) - 2T_3(x),$$

where $\gamma = 0.57721566...$ is Euler's constant. The example corresponds to $n = 3$ at line 1. Lines 2 and 3 are replaced by ExpansionCoeff $= [e, 2\pi, 2\gamma, -2]$;. The roots are approximately 1.44, -1.02 and -0.13. The two roots outside of $[-1, 1]$ due to their large condition numbers (estimated by Vcolsums).

In the algorithm, three user supplied external functions are called. The function setupChebyshev() determines the CGL points and the matrix that transforms function values to expansion coefficients.

```
function [cgl,CT] = setupChebyshev(n)
y = [0:n]*pi/n;
cgl = cos(y);
for i=0:n,
   CT(i+1,:) = cos(y*i);
end
pp = ones(n+1,1); pp(1) = 1/2; pp(n+1) = 1/2;
CT = diag(pp) * CT * diag(pp);
CT = CT * (2/n); %                    End of function setupChebyshev
```

The function problemRod evaluates user function at specified points in the domain. Here a problem associated with the vibration of an elastic rod is solved.

```
function functValues = problemRod(cgl)
[one, ncol] = size(cgl);
n = ncol-1;
first = cgl*3 + ones(1,n+1)*(3+1);
functValues = cos(first*pi) - sech(first*pi);
%                                         End of function problemRod
```

The vector ExpansionCoeff is the vector of coefficients in the collocation approximation by Chebyshev polynomials of degree up to n. The function evalCheb evaluates the Chebyshev polynomials at a specified set of points.

```
function V = evalCheb(degree_max,z)
% Input: vector of points, z, and the polynomial degree, degree_max.
% Output: Vandermonde matrix, m by degree_max + 1, V(j+1,k)=T_j(z_k)
[m,one] = size(z);
if m*degree_max >= 0,
   V(:,1) = ones(m,1);
   if degree_max >= 1,
      V(:,2) = z;
      if degree_max >= 2,
          index = find( log(abs(z))  >= 100/degree_max ); %  avoid
          si = size(index,1);                              % overflow
          if si > 0
             z(index) = ones(si,1)*exp(100/degree_max);
          end
          for i=2:degree_max,
             V(:,i+1) = V(:,i).*(2*z)  - V(:,i-1);
          end
      end
```

```
   end
else
   V = [];
end %                                             End of function evalCheb
```

In practice, five figures are recommended for verification purposes. Graphs of the expansion coefficients of the function and its derivative versus the indices are useful for assessing convergence. The exponential decay of the expansion coefficients indicates the convergence of the series to the transcendental equation. A plot of the cubic spline interpolants at the CGL points to the function and its derivative over the interval $[-1, 1]$ is useful for detecting clusters of roots. Later the (real parts of the) roots and the (real parts of the) derivative values at the roots may be overlaid onto the respective graphs. In selecting the eigenvalues that approximate roots of the polynomial or transcendental equation, it is helpful to compare the distribution of the eigenvalues and the row sums of the generalized Vandermonde matrix. The roots in $[-1, 1]$ and the other roots appear as two sets that are easier to see than to quantify. Finally one must check the transcendental equation residuals at the selected eigenvalues.

5. Examples. We apply our rootfinding technique to Chebyshev expansions arising in the solution of some transcendental equations. Two representative applications of the algorithm are discussed followed by two more challenging applications. For a transcendental equation $\psi(\xi) = 0$ with clustered or multiple roots, an added issue is that roots of $p_n(x)$ that are not near roots of $\psi(x)$ appear among the approximate solutions of $\psi(\xi) = 0$. Such problems demonstrate the importance of monitoring $\psi'(x)$. Examples 1 and 2 concern problems with well separated simple roots. For polynomials with multiple roots, $p_n(x)$ may very accurately approximates $\psi(x)$, and still have spurious roots near to roots of $\psi(x)$, as will be shown in Example 3. Lastly numerical experiments on computing the eigenvalues of $\mathbf{B}_n$ are discussed.

The methodology of the experiments is as follows. The number of function evaluations is carefully minimized. The degree of the polynomial n is doubled. In fact n is a power of 2 the the Examples except Example 2. Although doubling the polynomial order with a spectral method is over-kill, it is done here for two reasons. First, if the function is evaluated at n points, it is possible to reuse the previous $n/2$ function values [6], carefully minimizing the number of function evaluations. Second, we wish to illustrate the properties of the nonstandard companion matrices for all values of n, not just special values. We find that using excessively large values of n, up to 2^{10}, does not effect the accuracy of the approximate roots in $[-1, 1]$. For transcendental equations, we discuss at length the minimal values of n for which the series is (almost) converged.

Example 1 concerns the transverse vibrations (u) of a homogeneous rod of length π with both ends free (u'' and u''' vanish at endpoints). The first six flexible modes are found by solving the secular equation $\cos(\pi x) - \operatorname{sech}(\pi x)$ in the interval $1 \leq x \leq 7$ [9] page 296. For $n = 16$ or $n = 32$ approximates all six roots to within 5 or 14 significant digits respectively.

Example 2 reproduces results from [6]. The roots of Bessel's function of the first kind, $J_\nu(\xi) = 0$, are computed without doubling the polynomial degree. In the first numerical experiments $J_0(x) = 0$ is solved over three intervals, $[0, w]$, for $w = 20{,}60$ and 180; $J_0(x)$ has 6, 19 and 57 roots in the respective intervals. The computed roots J_0 are compared to the roots computed by a stable algorithm. Here we can exactly reproduce the results of [6].

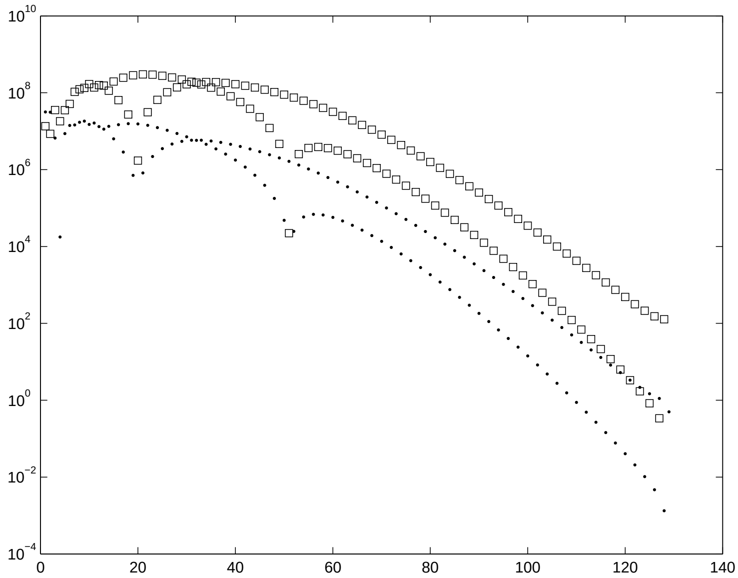

FIG. 5.1. *The function $\psi(x)$ of Example 3 is approximated by a Chebyshev series $p(x)$ of degree 128. The absolute values of the expansion coefficients of $p(x)$ (.) and $p'(x)$ (square) are plotted on a logarithmic scale.*

Example 3 originating in [21] is the nonlinear eigenvalue problem for $T(\lambda) = \lambda^2 B^{(2)} + (e^{\lambda} - 1)B^{(1)} - B^{(0)}$. The transcendental function here is that the determinant of $T(\lambda)$. The determinant is evaluated by factoring $T(\lambda)$ (not by Cramer's Rule). A large interval is chosen to test the numerical stability of of the rootfinding algorithm.

The problem is an example of an over damped system. Each $B^{(i)}$ is symmetric positive definite. At $\lambda = -\infty$, $\lambda = 0$ and $\lambda = +\infty$ the normalized matrix $T(\lambda)/\|T(\lambda)\|$ is positive definite, negative definite, and positive definite. The matrices are 8 by 8 with $B^{(0)} = 100I_8$, and for $0 \le i, j < 8$,

$$B^{(1)}_{i,j} = (i+1)(j+1)(9 - \max(i+1, j+1)) \quad \text{and} \quad B^{(2)}_{i,j} = 8\delta_{i,j} - 1/(i+j+2).$$

Six roots are clustered near to -3.7, with average absolute gap .1. As is carefully documented in [4], the the exponential growth of $\det T(\lambda)$ as $\lambda \to +\infty$ impedes resolution on certain intervals. The problem illustrates the rewards for choosing a suitable interval. On the interval $[-8, 8]$ in double precision arithmetic the roots of the polynomial approximation of $\det T(\lambda)$ poorly approximate the solutions, but in the interval $[-8, 4]$, the roots of the the polynomial approximation converge rapidly to the solutions.

Another approach, pursued here, concerns the alternative scaled problem $\psi(\lambda) = \det(\ T(\Lambda)/\sigma(\lambda)\ )$ for

$$\sigma(\lambda) = (\det B^{(0)})^{1/8} + (\det B^{(1)})^{1/8}(e^{\lambda} - 1) + (\det B^{(2)})^{1/8}\lambda^2.$$

The interval $[-10, 10]$ containing all 16 of the roots is used. We will discuss in detail the results obtained using a Chebyshev series expansions of degree 128 (see Figure 5.1). Of the 128 eigenvalues, 108 are discarded, and 20 are potential solutions (see Figure 5.2).

Inspection of the graphs of the $\psi(x)$ and $\psi'(x)$ shown in Figures 5.3 and 5.4 respectively is helpful. In a large neighborhood of the root cluster around -3.7, there

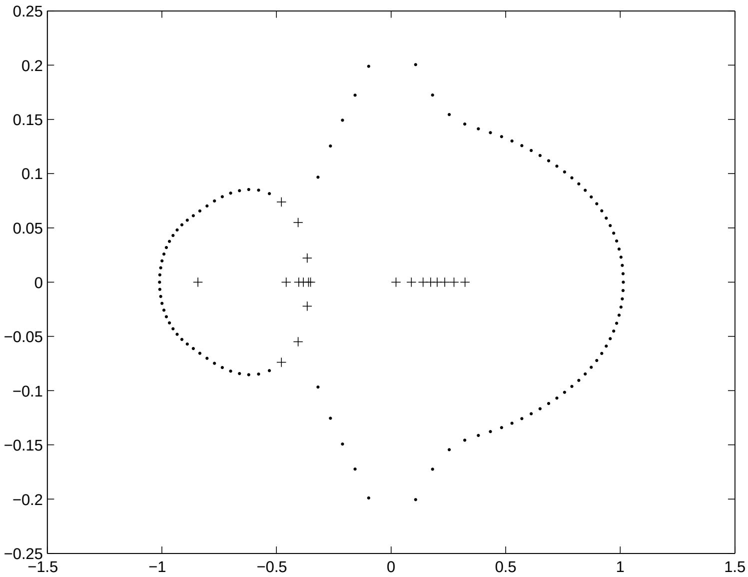

FIG. 5.2. *In Example 3 using a Chebyshev series of degree* 128 *results in a companion matrix* $\mathbf{B}_{128}$*. The figure displays the eigenvalues of* $\mathbf{B}_{128}$ *in the complex plane. A* $\cdot$ *indicates each of the* 108 *discarded eigenvalues, and a* $+$ *indicates each of the* 20 *potential roots.*

holds $|\psi| < 10^{-9}$ and $|\psi'| < 10^{-5}$. A degree 128 polynomial is insufficient to resolve each root in the cluster. On the other hand, with a degree 256 polynomial, 240 eigenvalues are discarded. The remaining 16 eigenvalues approximate the 16 roots, each with residual norms below 10^{-13}.

Before closing, we make an observation about accelerating the convergence of the QR algorithm for computing the eigenvalues of $\mathbf{B}_n$ arising from the solution of transcendental equations. Loosely speaking, a matrix is graded (by diagonal) if the norms of the diagonals increase geometrically. Note that $\mathbf{B}_n$ is graded by diagonal. We have observed that the computed Schur form of $\mathbf{B}_n$ is also graded. Careful examination of the QR iterates (say with zero shifts) from $\mathbf{B}_n$ (without balancing) indicates that along with the accuracy of the computed eigenvalues, the graded structure is also lost. Better results are obtained using the matrix $\mathbf{M}_n\mathbf{B}_n^T\mathbf{M}_n^T$ determined using the anti-diagonal matrix $\mathbf{M}_n = [\mu_{i,j}]_{0\le i,j<n}$ with $\mu_{i,j} = \delta_{i,n-i-1}$. The matrix $\mathbf{M}_n\mathbf{B}_n^T\mathbf{M}_n^T$ is similar to $\mathbf{B}_n$ and inherits its unreduced upper Hessenberg and graded structure. We observe that the QR iteration applied to $\mathbf{M}_n\mathbf{B}_n^T\mathbf{M}_n^T$ (without balancing) preserves the graded structure of $\mathbf{B}_n$ and converges in many fewer iterations.

6. Conclusion. We have shown how to find the roots of a degree n polynomial $p(x)$ expressed in terms of orthogonal polynomials. In particular, we have shown that these roots are the eigenvalues of a nonstandard companion matrix $\mathbf{B}_n$. This companion matrix gets infinitely large as the highest order coefficient γ_n in our orthogonal expansion goes to zero. However, we have analyzed the numerical stability of this algorithm for Jacobi polynomials and found that it has good numerical stability properties as long as we are only interested in roots in the interval $[-1, 1]$. This makes the algorithm particularly suited for finding the roots of transcendental equations.

We have presented an algorithm for finding the roots of a scalar transcendental equation by expressing it in terms of orthogonal polynomials, and using the companion

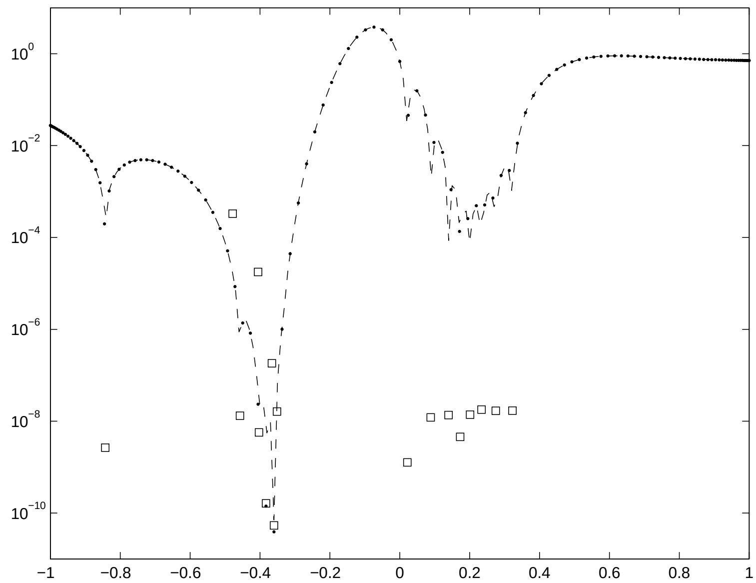

FIG. 5.3. *The value of $|p(x)|$ for $-1 \leq x \leq 1$ is shown on a logarithmic scale for the degree* 128 *Chebyshev series approximation of the function $\psi(x)$ of Example 3. The function values at CGL points (.), a spline interpolant to the CGL points (dashed line), and the residuals at the* 20 *potential roots (square) are each presented. A complex root ξ is displayed at $x = \Re(\xi)$. Due to this discrepancy, such function values appear above the spline interpolant in* $[-1, 1]$.

matrix $\mathbf{B}_n$. We have given several numerical examples that illustrate the stability of this algorithm. For a more detailed summary, see §1.1.

Acknowledgments. It is a pleasure to acknowledge the assistance of the referees Prof. L.N. Trefethen and the anonymous referee in the preparation of the document. The authors wish to thank Prof. Holly Dison for assistance in performing the numerical experiments and Prof. Ming Gu who we heard had used the Chebyshev companion matrix, though he has not published anything on any nonstandard companion matrix.

REFERENCES

[1] G.S. Ammar, D. Calvetti, W.B. Gragg, and L. Reichel. Polynomial zerofinders based on Szegö polynomials. *J. Comput. Appl. Math.*, 127(1-2):1–16, January 2001.

[2] Z. Battles and L. N. Trefethen. An extension of matlab to continuous functions and operators. *SIAM J. Sci. Comput.*, 2004.

[3] C. Bernardi and Y. Maday. *Approximations Spectrals de Problèms aux Limites Elliptiques*, volume 10 of *Collection Mathématiques et Applications*. Springer, Paris, 1992.

[4] J.P. Boyd. A Chebyshev polynomial interval-searching method ("Lanczos economization") for solving a nonlinear equation with application to the nonlinear eigenvalue problem. *J. Comp. Phys.*, 118:1–8, 1995.

[5] J.P. Boyd. *Chebyshev and Fourier Spectral Methods*. Dover, New York, 2nd edition, 2001.

[6] J.P. Boyd. Computing zeros on a real interval through Chebyshev expansions and polynomial root finding. *SIAM J. Numer. Anal.*, 40(5):1666–1682, 2002.

[7] J.P. Boyd. Computing real roots of a polynomial in chebyshev series form through subdivision. Technical Report OP145, UM, US, 2004. Submitted to Applied Numerical Math.

[8] J.P. Boyd. Computing real roots of a polynomial in chebyshev series form through subdivision with linear testing and cubic solves. Technical Report OP145, UM, US, 2004. Submitted to Applied Math. and Computation.

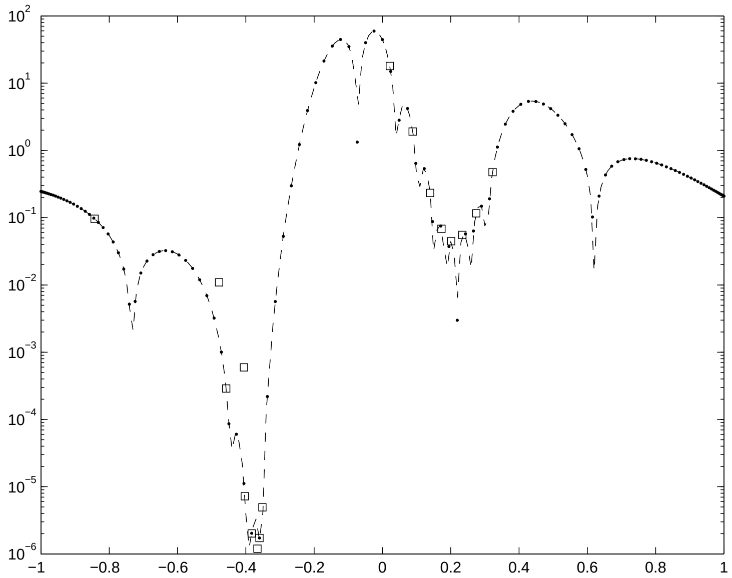

FIG. 5.4. *Example 3, The absolute value the derivative* $|p'(x)|$ *for* $-1 \leq x \leq 1$ *is shown on a logarithmic scale for the degree* 128 *Chebyshev series approximation of the function* $\psi(x)$ *of Example 3. The function values at CGL points (.), a spline interpolant to to the CGL points (dashed line), and the derivative values at the* 20 *potential roots (square) are all shown. Complex roots are displayed at the real part of the root, and due to this discrepancy, such function values are differ from the spline interpolant.*

[9] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 1. Interscience Publishers, NY, 1st english edition, 1937.
[10] S. Fortune. An iterated eigenvalue algorithm for approximating roots of univariate polynomials. *Journal of Symbolic Computation*, 33:627–646, 2002.
[11] W. Gautschi. The condition of orthogonal polynomials. *Math. Comp.*, 26(120):923–924, 1972.
[12] W. Gautschi. The condition of algebraic equations. *Numer. Math.*, 21:405–424, 1973.
[13] W. Gautschi. The condition of polynomials in power form. *Math. Comp.*, 33(145):343–352, 1979.
[14] W. Gautschi. Orthogonal polynomials: applications and computation. *Acta Numerica*, 5:45–120, 1996.
[15] G. H. Golub and C. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, 2nd edition, 1989.
[16] D. Gottlieb and S.A. Orszag. *Numerical Analysis of Spectral Methods: Theory and Applications.* SIAM, Philadelphia, PA, USA, 2nd edition, 1989.
[17] N.J. Higham. *Accuracy and Stability of Numerical Algorithms.* SIAM, 1st edition, 1996.
[18] E. Hille. *Analytic Function Theory*, volume II. Ginn and Company, Boston, 1962.
[19] H. Hochstadt. *The Functions of Mathematical Physics.* Dover, 1986.
[20] G. F. Jónsson and S. Vavasis. Solving polynomials with small leading coefficients. *SIAM J. Matrix Anal. Applic.*, 26:400–414, 2004.
[21] K.P.Hadeler. Mehrparametrige und nichtlineare eigenwertaufgaben. *Arch. Rational Mech. Anal.*, 27:306–328, 1967.
[22] H.J. Stetter. *Numerical Polynomial Algebra.* SIAM, 1st edition, 2004.
[23] G. Szegö. *Orthogonal Polynomials.* American Mathematical Society, 4th edition, 1975.
[24] K.-C. Toh and L.N. Trefethen. Pseudozeros of polynomials and pseudospectra of companion matrices. *Numer. Math.*, 68(3):403–425, 1994.